https://doi.org/10.31891/2219-9365-2025-81-53
УДК 004.4

МІХАЛЕВСЬКИЙ Віталій
Хмельницький національний університет
https://orcid.org/0000-0002-8197-8005
e-mail: cezar_mv@ukr.net
СКРИПНИК Тетяна
Хмельницький національний університет
https://orcid.org/0000-0002-8531-5348
e-mail: tkskripnik1970@gmail.com
МЕДВЕДЧУК Неля
Хмельницький національний університет
https://orcid.org/0000-0003-1337-2121
e-mail: medvedchuk.nelya@gmail.com
ВОЗНЮК Леонід
Хмельницький національний університет
https://orcid.org/0009-0002-1152-3192
e-mail: leonid.voznuk@gmail.com

ДИНАМІЧНІ МНОЖИНИ ДАНИХ ДЛЯ ЗАСОБІВ МАШИННОГО НАВЧАННЯ В ІТ-ПРОЄКТІ ПРОГНОЗУВАННЯ ЕРОЗІЇ ҐРУНТІВ

В статті розглянуто використання динамічних множин даних для засобів машинного навчання в ІТ-проєкті прогнозування ерозії ґрунтів. Суть підходу полягає у використанні алгоритмів для аналізу великих обсягів динамічних даних про ґрунтові, топографічні, метеорологічні та рослинні параметри з метою виявлення закономірностей, що дозволяють прогнозувати ризики ерозії. Використання методів машинного навчання для прогнозування ерозії ґрунтів є перспективним напрямком, оскільки дозволяє автоматизувати процеси обробки великих масивів даних і отримувати прогнози, що є значно точнішими та швидшими порівняно з традиційними методами. Завдяки розвитку новітніх технологій у галузі машинного навчання, зокрема глибинного навчання та алгоритмів класифікації, можна створювати високоточні прогнози, що дозволять більш ефективно зберігати ґрунти, попереджати ерозію та мінімізувати її негативні наслідки. ІТ-проєкт створення точних моделей на множинах динамічної різнорідної інформації забезпечує оперативне, обгрунтоване і відповідальне прийняття рішень для управління земельними ресурсами та запобігання деградації ґрунтів.

Ключові слова: ерозія ґрунту, машинне навчання, прогнозування, динамічна множина, ІТ-проєкт.

MIKHALEVSKYI Vitaliy, SKRYPNYK Tetiana, MEDVEDCHUK Nelya, VOZNIUK Leonid
Khmelnytsky National University

DYNAMIC DATASETS FOR MACHINE LEARNING IN THE SOIL EROSION FORECASTING IT PROJECT

The article discusses the use of dynamic datasets for machine learning tools in an IT project for soil erosion forecasting. The essence of the approach is to use algorithms to analyze large volumes of dynamic data on soil, topographic, meteorological and vegetation parameters in order to identify patterns that allow predicting erosion risks. The use of machine learning methods for soil erosion forecasting is a promising direction, as it allows automating the processes of processing large data sets and obtaining forecasts that are much more accurate and faster compared to traditional methods. Thanks to the development of the latest technologies in the field of machine learning, in particular deep learning and classification algorithms, it is possible to create highly accurate forecasts that will allow more effectively preserving soils, preventing erosion and minimizing its negative consequences. The IT project for creating accurate models on sets of dynamic heterogeneous information provides prompt, informed and responsible decision-making for land resource management and preventing soil degradation. It also leads to a deterioration in the water balance, since soil erosion is often accompanied by pollution of water resources, which reduces their quality and suitability for use. In this regard, there is an urgent need to create accurate erosion forecasting models that will help identify areas of increased risk and promote effective management of natural resources. Thanks to the development of the latest technologies in the field of machine learning, in particular deep learning and classification algorithms, it is possible to create highly accurate forecasts that will allow you to more effectively preserve soils, prevent erosion and minimize its negative consequences. This approach allows for the creation of accurate models that help in making informed decisions for land management and preventing soil degradation.

Keywords: soil erosion, machine learning, forecasting, dynamic set, IT project.

ПОСТАНОВКА ПРОБЛЕМИ У ЗАГАЛЬНОМУ ВИГЛЯДІ ТА ЇЇ ЗВ'ЯЗОК ІЗ ВАЖЛИВИМИ НАУКОВИМИ ЧИ ПРАКТИЧНИМИ ЗАВДАННЯМИ

Ерозія ґрунту – це складне явище, що виникає внаслідок дії природних і антропогенних факторів, таких як вітер, вода, сільськогосподарська діяльність, вирубка лісів та урбанізація. За оцінками, близько 23% земної поверхні вже постраждали від різних форм ерозії, що призводить до щорічних втрат до 28 мільярдів тон родючого шару ґрунту. Особливо гостро це питання постає у регіонах із засушливим кліматом, де ґрунти менш стійкі до вітрової ерозії, а також у районах із інтенсивним землеробством. Однією з найпоширеніших

форм є водна ерозія, яка охоплює процеси поверхневого змиву ґрунту дощовими водами, утворення ярів і ерозійних рівчаків. Її інтенсивність залежить від таких факторів, як нахил місцевості, кількість і частота опадів, структура ґрунту та наявність рослинного покриву. Наприклад, у зонах з нахилом понад 12° інтенсивність змиву ґрунту збільшується майже на 17%. У сільськогосподарських районах з активним застосуванням оранки на схилах ці показники можуть бути ще вищими, досягаючи 34% втрат ґрунту за рік. Вітрова ерозія, або дефляція, є характерною для регіонів із сухим кліматом і низькою рослинністю. Вітер переносить частинки ґрунту на значні відстані, спричиняючи утворення пилових бур і знижуючи родючість. Відомо, що щільність часток ґрунту впливає на його здатність протистояти вітровій ерозії: піщані ґрунти втрачають до 37% родючого шару за рік у несприятливих умовах. Зменшення площі лісових масивів, особливо у степових і напівпустельних зонах, значно погіршує ситуацію, оскільки лісові екосистеми відіграють роль природного бар'єра проти вітрових потоків [1, 2].

Сучасні дослідження зосереджені на використанні новітніх технологій, таких як засоби машинного навчання, для прогнозування ерозійних процесів. Вони дозволяють створювати точні моделі, що враховують широкий спектр факторів – від кліматичних змін до структури земельного покриву та антропогенних впливів. Наприклад, використання супутникових знімків у поєднанні з алгоритмами глибокого навчання дає змогу з точністю до 91% передбачати зони високого ризику ерозії, що є значним проривом у запобіганні деградації ґрунтів [3]. У результаті аналіз предметної області ерозії ґрунту дозволяє виявити ключові проблеми, які необхідно вирішити для забезпечення сталого землекористування. Ключовим аспектом аналізу ерозії ґрунту є її вплив на сільське господарство, яке значною мірою залежить від родючості ґрунтів. Деградація орних земель через ерозію призводить до зниження врожайності сільськогосподарських культур на 20–40% у найбільш постраждалих регіонах. Наприклад, втрати родючого шару на глибину всього 2–3 см можуть зменшити врожайність пшениці на 15–30%, оскільки зменшується здатність ґрунту утримувати вологу та поживні речовини. Це створює серйозні виклики для забезпечення продовольчої безпеки, особливо в умовах зростання світового населення та попиту на продукти харчування.

Аналіз предметної області ерозії ґрунту демонструє її багатогранний характер і необхідність інтегрованого підходу до вирішення цієї глобальної проблеми. Використання сучасних наукових підходів, поєднаних із практичними рішеннями на місцевому рівні, може суттєво зменшити наслідки ерозії, забезпечуючи стале майбутнє для сільського господарства, довкілля та суспільства загалом.

Сучасні дослідження зосереджені на використанні інноваційних технологій для картографування й аналізу ерозії. Машинне навчання дозволяє створювати високоточні карти ризику ерозії з урахуванням локальних кліматичних і географічних умов. Наприклад, алгоритми кластеризації застосовуються для аналізу геопросторових даних, що включають інформацію про нахил поверхні, вологість, текстуру ґрунту і щільність рослинності. В Україні, зокрема, дослідницькі проекти, які використовують супутникові дані Sentinel-2, показали можливість створення прогнозів із точністю до 93%, що дає змогу розробляти ефективні стратегії для запобігання ерозії. Подальше вдосконалення методів прогнозування ерозії грунту є важливим фактором економічного розвитку та екологічної рівноваги навколишнього середовища. Необхідно розробити і дослідити метод прогнозування ерозії грунту засобами машинного навчання для управління природними ресурсами.

АНАЛІЗ ДОСЛІДЖЕНЬ ТА ПУБЛІКАЦІЙ

Інформаційні технології на сьогоднішній день широко застосовуються для дослідження і прогнозування ерозії ґрунтів. При цьому важливим фактором є наукове обґрунтування існуючих проблем та шляхів їх вирішення. Вхідними даними для методів наукового вивчення процесів ерозії грунту є географічні фактори, які детально аналізуються багатьма авторами. Наприклад, в роботах Волощука М. Д. [4] досліджуються особливості ерозії ґрунтів Південно-західної частини України та Республіки Молдова за географічними факторами. Автор приділяє значну увагу факторам, таким як рельєф, кліматичні умови та типи ґрунтів, які впливають на інтенсивність ерозійних процесів. Важливим результатом є аналіз просторового розподілу ерозійних ділянок, що дозволяє виявити критичні зони для застосування захисних заходів. Також можуть застосовуватися інші сучасні технології, такі як геоінформаційні системи (ГІС) та дистанційне зондування Землі. Ці інструменти дозволяють виконувати високоточний аналіз просторових даних, що сприяє створенню інтегрованих підходів до оцінки ризиків ерозії та планування природоохоронних заходів.

Застосування методів машинного навчання розглядається як один із основних методів дослідження та прогнозування ерозії. В роботах Грабовського А. О. та Ляшенка В. Р. [5] розглядається можливість прогнозування ерозії ґрунтів на основі супутникових знімків. Автори підкреслюють, що використання нейронних мереж і різноманітних алгоритмів забезпечує високу точність прогнозів, що є критично важливим для ефективного управління земельними ресурсами. Крім того, такі методи дозволяють значно знизити час і витрати на моніторинг ерозійних процесів. Методи машинного навчання для оцінки ризиків ерозії ґрунтів в умовах змінного клімату аналізуються в статті Іванова О. В. [6]. Він порівнює різні підходи, включаючи методи класифікації та регресії для обробки просторових даних, що дозволяє ефективно прогнозувати зміни в стані ґрунтів. Також підкреслюється важливість врахування кліматичних і географічних факторів для

створення точних прогнозних моделей. Задачу використання глибинного навчання для прогнозування ерозії ґрунтів розглядають у своїх працях науковці Бойко П. Д., Черненко Л. О. [7]. У своїх дослідженнях вони розглядають можливості застосування глибинних нейронних мереж для обробки супутникових зображень та визначення зон з високим ризиком ерозії. Автори також зазначають, що такий підхід дозволяє значно покращити точність прогнозів і виявлення потенційно небезпечних територій для своєчасного реагування. Івахненко О. М., Івахненко Е. Н. [8] у своїй роботі розглядають ерозію ґрунтів як проблему сталого аграрного землекористування. Автори пропонують практичні рекомендації щодо інтеграції природоохоронних заходів у сільське господарство, таких як організація сівозмін, створення лісосмуг та використання сучасних методів моніторингу ерозії. Ці підходи спрямовані на запобігання деградації ґрунтів, що є ключовою умовою для забезпечення довгострокової стійкості аграрного сектору.

ФОРМУЛЮВАННЯ ЦІЛЕЙ СТАТТІ

Метою роботи є дослідження динамічних множин даних для засобів машинного навчання в ІТ-проєкті прогнозування ерозії грунтів та розробка теоретичних і практичних напрацювань для оцінки ризиків ерозії ґрунтів та визначення зон з високим ризиком ерозії.

ВИКЛАД ОСНОВНОГО МАТЕРІАЛУ

Концепція проєкту розробки прогнозування ерозії ґрунту засобами машинного навчання базується на використанні сучасних інструментів аналізу великих даних, що дозволяє враховувати багатовимірний характер цього явища. Основна ідея методу полягає в тому, щоб розробити модель, здатну точно передбачати ризики ерозії на основі обробки значної кількості різнорідної інформації, включаючи кліматичні дані, топографічні особливості, характеристики рослинного покриву та інтенсивність антропогенного впливу. Це дозволяє не лише ідентифікувати зони підвищеного ризику, але й розробити рекомендації для ефективного управління земельними ресурсами. Ключовим елементом концепції є інтеграція даних із різних джерел, таких як супутникові зображення, метеорологічні спостереження та польові вимірювання. Наприклад, дані про кліматичні умови включають середньорічну кількість опадів, інтенсивність зливових дощів та частоту замерзання й відтавання ґрунту. Ці фактори впливають на водну ерозію, яка є основною причиною втрати верхнього родючого шару ґрунту. Для визначення інтенсивності ерозійних процесів також використовуються показники ухилу території, висоти над рівнем моря та геологічної структури ґрунту. У регіонах із крутими схилами (понад 15°) втрати ґрунту можуть досягати до 30 тон на гектар на рік, тоді як у рівнинних місцевостях ці показники значно нижчі [4].

Другим важливим компонентом є використання алгоритмів машинного навчання для побудови моделі прогнозування. Алгоритми, такі як лінійна регресія, дерева рішень або нейронні мережі, дозволяють знаходити складні нелінійні залежності між вхідними параметрами та результатами ерозійних процесів. Наприклад, нейронні мережі мають високу здатність до генералізації і можуть враховувати вплив малопомітних факторів, таких як сезонні коливання температур або тимчасові зміни у структурі ґрунту. Перед початком моделювання всі дані проходять етап попередньої обробки: видалення шумів, заповнення пропусків і нормалізація значень. Це забезпечує більшу точність та надійність прогнозів, оскільки відсутність систематичних помилок на початковому етапі є критичним для ефективності моделі [5, 6].

Проаналізувавши чотири відомих методи машинного навчання (дерево рішень, регресія, ансамблевий, нейроні мережі) вибираємо для застосування в нашому методі прогнозування ерозії грунтів метод Random forest. Особливе місце в концепції методу прогнозування займає визначення ключових характеристик, які найбільше впливають на ерозійні процеси. До них належать нахил місцевості, густота рослинного покриву, середньорічна кількість опадів, тип ґрунту та рівень антропогенного впливу. Відсутність рослинного покриву збільшує ризик ерозії у 5-10 разів у порівнянні з територіями, вкритими густою рослинністю. Аналіз цих показників дозволяє не лише створити ефективну модель, але й визначити конкретні заходи для зниження ризиків, такі як відновлення рослинного покриву чи впровадження агротехнічних практик, орієнтованих на попередження ерозії. Завершальним етапом концепції є інтеграція створеної моделі у практичні системи управління. Це включає розробку геоінформаційних систем (ГІС), що надають користувачам інструменти для візуалізації ризиків ерозії в режимі реального часу. Наприклад, фермери можуть отримувати попередження про потенційні ризики на основі прогнозованих кліматичних умов і планувати сільськогосподарські заходи таким чином, щоб мінімізувати втрати ґрунту. У поєднанні з політиками сталого розвитку це дозволяє забезпечити довгострокову продуктивність земель і зберегти екологічну рівновагу [3].

Рис. 1. Узагальнена схема методу прогнозування ерозії ґрунту засобами машинного навчання

Процес прогнозування ерозії ґрунту є багатокомпонентною задачею, яка включає в себе аналіз численних факторів, що впливають на ерозійні процеси і їх взаємодію. Інформаційна модель методу є основою для розробки ефективної системи прогнозування. Ця модель має включати сукупність структурованих даних, алгоритмів машинного навчання та параметрів, що забезпечують інтеграцію і аналіз різнорідної інформації про стан ґрунту, кліматичні умови та антропогенні впливи. Модель будується на основі даних, зібраних з різних джерел: географічних інформаційних систем (ГІС), метеорологічних станцій, супутникових знімків, польових спостережень та історичних записів. Ці дані включають такі показники, як тип ґрунту, нахил місцевості, опади, швидкість вітру, щільність рослинного покриву та агротехнічні заходи, які застосовуються на ділянці. Усі ці параметри є змінними, що впливають на ерозійні процеси, і повинні бути враховані у моделі [1, 5].

Алгоритми машинного навчання, які використовуються у моделі, обираються відповідно до поставленої задачі та характеру даних. Найпоширенішими підходами є регресійні методи, дерева рішень, ансамблеві методи, нейронні мережі та методи кластеризації. Вибір методу залежить від того, чи задача є задачею класифікації (наприклад, визначення рівня ризику ерозії) або регресії (оцінка кількісного рівня втрати ґрунту).

Узагальнена схема методу прогнозування ерозії ґрунту засобами машинного навчання передбачає кілька ключових етапів, кожен з яких має суттєве значення для забезпечення точності та надійності кінцевих результатів. Першим кроком є збір даних, що включає отримання інформації про топографію, кліматичні умови, рослинний покрив, а також інші фактори, що впливають на процеси ерозії ґрунту. Ці дані можуть бути зібрані з різних джерел, включаючи супутникові знімки, метеорологічні станції та польові дослідження [1, 6].

Після збору даних проводиться їх попередня обробка, яка включає очищення від шумів, заповнення пропусків та нормалізацію значень для подальшої обробки. Важливим етапом є вибір характеристик, що представляють собою ключові параметри, які найбільш суттєво впливають на ерозійні процеси. Від правильного вибору характеристик залежить ефективність подальшого моделювання та прогнозування.

Наступним етапом є розділення даних на тренувальний і тестовий набори, що дозволяє оцінити точність моделі та запобігти перенавчанню. Вибір моделі машинного навчання є критично важливим кроком, оскільки різні алгоритми можуть мати різну ефективність для конкретних задач. Серед можливих моделей можуть бути регресійні моделі, нейронні мережі, дерева рішень та інші. Тренування моделі проводиться на тренувальному наборі даних, де модель навчається виявляти залежності між обраними характеристиками та результатами ерозійних процесів. Після цього модель оцінюється на тестовому наборі, що дозволяє

перевірити її здатність до узагальнення і точності прогнозування на нових даних. У разі необхідності проводиться оптимізація моделі, яка включає налаштування гіперпараметрів для досягнення максимальної точності.

Оцінка ефективності методів прогнозування ерозії ґрунту є надзвичайно важливим етапом у наукових дослідженнях, адже дозволяє підтвердити точність і надійність розробленої моделі. Для цього необхідно застосувати комплексний підхід, що враховує статистичні, просторово-часові та екологічні аспекти проблеми. Основою ефективної оцінки є порівняння передбачених результатів із фактичними даними, які отримані внаслідок польових вимірювань або тривалих спостережень за процесами ерозії. Це дає змогу забезпечити об'єктивність і репрезентативність висновків. Одним із ключових етапів оцінки є вибір відповідних метрик для вимірювання точності моделі. Найчастіше використовуються такі показники, як середня абсолютна похибка (MAE), середньоквадратична похибка (RMSE) та коефіцієнт детермінації (R^2). Ці показники дозволяють визначити, наскільки точно модель прогнозує обсяг і просторовий розподіл ерозійних процесів. Розрахунок точності прогнозування слід проводити на основі розділення даних на навчальну та тестову вибірки. Такий підхід дозволяє оцінити, чи не стала модель занадто спеціалізованою для аналізу конкретного набору даних (явище оверфітингу) або навпаки, чи не є вона недостатньо адаптованою (андерфітинг). Крім того, для отримання надійних результатів використовується техніка перехресної валідації (k-fold cross-validation), яка забезпечує рівномірний розподіл даних між навчанням і тестуванням.

Для оцінки точності та якості прогнозування ерозії ґрунту застосовуються метрики, які дозволяють кількісно визначити розбіжності між прогнозованими моделлю значеннями та реальними даними. Основні формули включають:

1. Середня абсолютна похибка (Mean Absolute Error, MAE):

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \widehat{y_i}|, \tag{1}$$

де y_i — фактичне значення, $\widehat{y_i}$ — прогнозоване значення, n — кількість спостережень. Цей показник визначає середню величину абсолютних відхилень прогнозу моделі від реальних значень.

2. Середньоквадратична похибка (Root Mean Squared Error, RMSE):

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{y_i})^2}, \tag{2}$$

RMSE підкреслює великі помилки сильніше, оскільки використовує квадрат відхилення, що дозволяє оцінити моделі з точки зору великих аномалій.

3. Коефіцієнт детермінації (R-squared, R^2):

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \widehat{y_i})^2}{\sum_{i=1}^{n}\left(y_i - \underline{y}\right)^2}, \tag{3}$$

де $\underline{y}$ — середнє значення фактичних даних. Цей коефіцієнт показує частку варіації залежної змінної, яка пояснюється моделлю.

4. Середня відносна похибка (Mean Relative Error, MRE):

$$MRE = \frac{1}{n}\sum_{i=1}^{n}\frac{|y_i - \widehat{y_i}|}{y_i} \cdot 100\%, \tag{4}$$

MRE дозволяє оцінити точність прогнозу у відсотках, що важливо для розуміння відносної похибки моделі.

Для перевірки узгодженості моделі з реальними даними рекомендується проводити крос-валідацію (k-fold cross-validation), де вибірка даних розбивається на k підмножин, і модель тренується $k-1$ разів з оцінкою на залишковій підмножині. Цей підхід дозволяє отримати середній показник точності моделі.

Додатково, для просторового аналізу точності використовуються показники схожості ерозійних карт, наприклад, коефіцієнт Джаккара:

$$J = \frac{|A \cap B|}{|A \cup B|}, \tag{5}$$

де A та B — множини пікселів передбачених і фактичних зон ерозії.

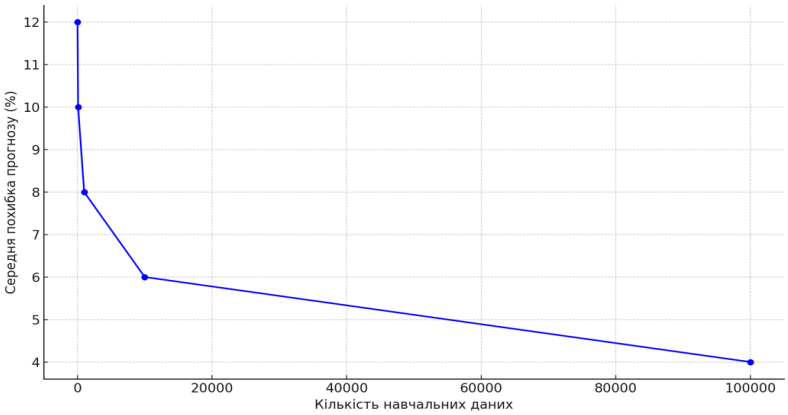

Рис. 2. Залежність точності прогнозу від кількості навчальних даних

Не менш важливим є просторовий аналіз точності прогнозування. Ерозія ґрунту є природним процесом, що має складну географічну структуру, тому ефективність моделі повинна оцінюватися не лише за статистичними показниками, але й з урахуванням просторової кореляції. Наприклад, моделі, що добре прогнозують ерозію в рівнинних районах, можуть демонструвати низьку точність у гірських регіонах. Застосування картографічних методів аналізу, таких як порівняння прогнозованих і фактичних ерозійних карт, є важливим інструментом оцінки.

Для підготовки навчальної множини даних збираємо різноманітну інформацію, яка охоплює топографічні, кліматичні, ґрунтові, біологічні та антропогенні фактори. На рисунку 2 представлено залежність точності прогнозу від кількості даних.

Набір даних для навчання моделі прогнозування ерозії ґрунту засобами машинного навчання включає конкретні значення, які відображають вплив природних і антропогенних умов на ерозійні процеси. У проведеному дослідженні використовувалися наступні показники.

Топографічні показники включають висоту над рівнем моря, нахил поверхні та експозицію схилу. Висота над рівнем моря для досліджуваної ділянки становить 350 метрів. Нахил поверхні вимірюється у градусах і дорівнює 25 градусам, що вказує на середній рівень крутизни. Експозиція схилу визначає напрямок нахилу і в даному випадку становить 135 градусів (південно-східний напрямок).

Кліматичні дані включають кількість опадів, середню температуру та швидкість вітру. За рік на досліджуваній території випадає 800 мм опадів. Середня температура становить 15° Цельсія, що є типовим для помірного клімату. Швидкість вітру на цій території складає 5 м/с, що вказує на помірні вітрові умови.

Дані про рослинний покрив включають тип рослинності, щільність покриття та сезонність. Тип рослинності представлений трав'яним покривом. Щільність покриття становить 75%, що забезпечує значну захисну функцію проти ерозії. Сезонність враховує, що рослинний покрив є густим протягом весняно-літнього періоду та зменшується до 50% у зимовий період.

Антропогенні фактори включають дані про сільськогосподарську діяльність, використання земель та наявність інфраструктури. На досліджуваній території проводиться інтенсивна оранка із застосуванням підсіву. Використання земель визначається як 80% сільськогосподарське, 10% урбанізоване і 10% природне. Інфраструктура включає 5 км доріг і 2 будівлі, що розташовані на території дослідження.

Ці конкретні значення даних використовуються для тренування моделей машинного навчання. Наприклад, висота над рівнем моря (350 м), нахил поверхні (25°), кількість опадів (800 мм), середня температура (15° Цельсія), швидкість вітру (5 м/с), щільність рослинного покриву (75%), інтенсивність сільськогосподарської діяльності (оранка з підсівом) та наявність інфраструктури (5 км доріг, 2 будівлі) є ключовими параметрами для моделювання ерозійних процесів.

Набір даних для тестування моделі прогнозування ерозії ґрунту засобами машинного навчання повинен включати конкретні значення, які відрізняються від тих, що використовувалися для навчання, але все ще представляють реальні умови, що впливають на ерозійні процеси.

Топографічні показники для тестової ділянки включають висоту над рівнем моря, нахил поверхні та експозицію схилу. Висота над рівнем моря становить 600 метрів. Нахил поверхні вимірюється у градусах і дорівнює 30°, що вказує на більш крутий схил порівняно з тренувальними даними. Експозиція схилу становить 180° (південний напрямок).

Кліматичні дані для тестової ділянки включають кількість опадів, середню температуру та швидкість вітру. За рік на цій території випадає 1000 мм опадів, що значно більше, ніж у тренувальному наборі. Середня температура становить 10° Цельсія, що є значно холоднішим порівняно з тренувальними даними. Швидкість вітру на цій території складає 8 м/с, що вказує на сильні вітрові умови. Дані про рослинний покрив для тестової ділянки включають тип рослинності, щільність покриття та сезонність. Тип рослинності

представлений лісовим покривом. Щільність покриття становить 90%, що є вищим, ніж у тренувальних даних, забезпечуючи кращий захист від ерозії. Сезонність враховує, що рослинний покрив залишається густим протягом усього року.

Антропогенні фактори для тестової ділянки включають дані про сільськогосподарську діяльність, використання земель та наявність інфраструктури. На цій території проводиться органічне землеробство, яке включає мінімальну механічну обробку ґрунту та використання природних добрив. Використання земель визначається як 50% сільськогосподарське, 40% природне і 10% урбанізоване. Інфраструктура включає 2 км доріг і 3 будівлі, що розташовані на території дослідження.

Ці конкретні значення даних використовуються для тестування моделей машинного навчання. Наприклад, висота над рівнем моря (600 м), нахил поверхні (30°), кількість опадів (1000 мм), середня температура (10° Цельсія), швидкість вітру (8 м/с), щільність рослинного покриву (90%), тип рослинності (лісовий покрив), інтенсивність сільськогосподарської діяльності (органічне землеробство) та наявність інфраструктури (2 км доріг, 3 будівлі) є ключовими параметрами для оцінки ефективності моделі.

Не всі змінні у прогнозування ерозії грунтів рівнозначні. Якщо візьмемо важливість всіх показників за 100%, то проаналізувавши їх питому вагу, можемо роботи попередні висновки про ерозію на основі кількох найважливіших факторів, що спрощує процес прогнозування. На рисунку 3 представлено графік важливості змінних у прогнозуванні ерозії ґрунту.

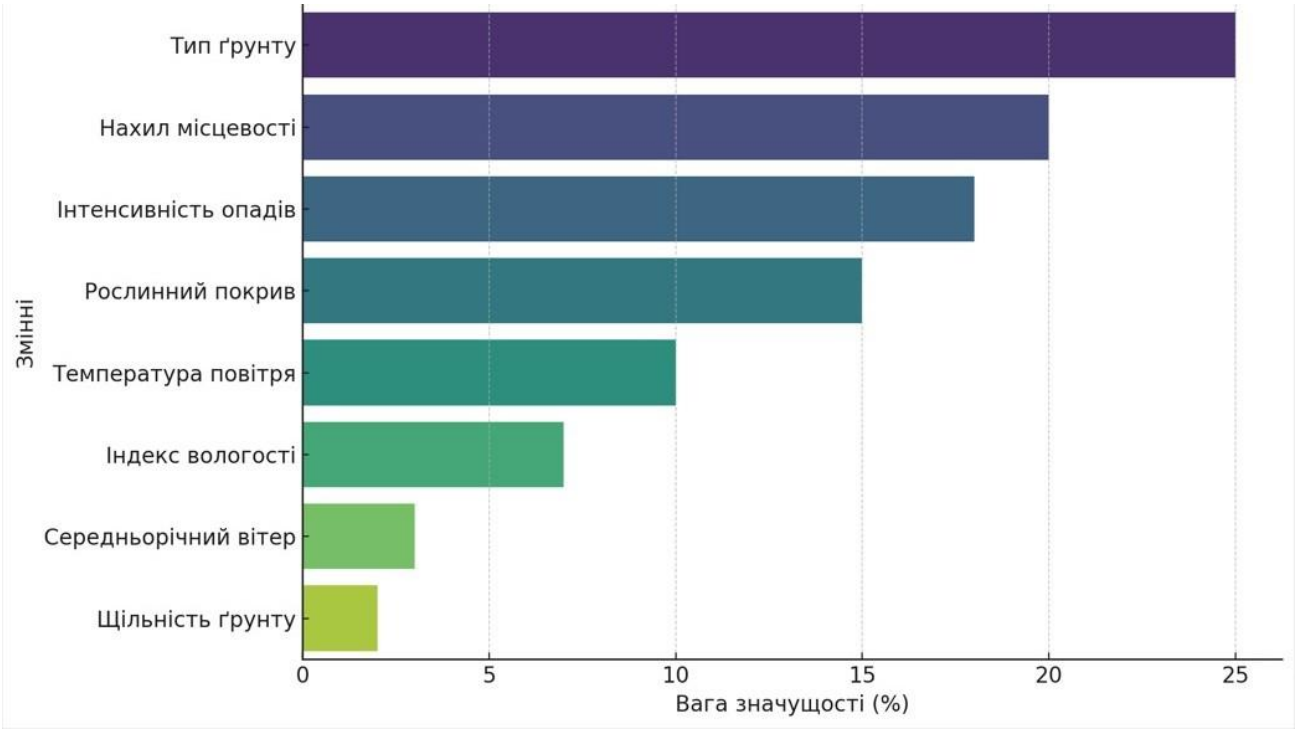

Рис. 3. Графік важливості змінних у прогнозуванні ерозії ґрунту

Найбільший вплив на точність прогнозу мають такі змінні, як тип ґрунту, нахил місцевості та інтенсивність опадів. Разом вони забезпечують понад 60% точності моделі. Параметри, пов'язані з мікрокліматом (температура, індекс вологості, вітер), також відіграють важливу роль, але їхній вплив менш значущий. Найменший вплив спостерігається від щільності ґрунту та середньорічної швидкості вітру. Це свідчить про те, що ці змінні можна використовувати як додаткові, але вони не є критичними для прогнозування. Графік важливості змінних наочно демонструє, що перші три фактори суттєво переважають інші. Такі висновки дозволяють спростити процес збору даних у реальних умовах, зосередившись на ключових параметрах.

Ще одним важливим дослідженням було тестування альтернативних моделей машинного навчання для оцінки їхньої продуктивності у прогнозуванні ерозії ґрунту. Одним з найважливіших аспектів вибору алгоритму є здатність моделі обробляти великі обсяги даних та враховувати складні взаємозв'язки між різними факторами, що впливають на ерозію. Серед багатьох існуючих алгоритмів машинного навчання найбільш перспективними для прогнозування ерозії ґрунту є регресійні моделі, методи опорних векторів, рішення дерев та нейронні мережі.

Регресійні моделі, такі як лінійна регресія, можуть бути використані для моделювання залежності ерозійних процесів від різних змінних, таких як кількість опадів, нахил місцевості, рослинний покрив та інші

кліматичні параметри. Проте, оскільки ерозія ґрунту є складним нелінійним процесом, прості регресійні моделі можуть бути недостатньо точними для прогнозування. Тому доцільно застосовувати поліноміальні регресії або регресійні моделі з регуляризацією, такі як гребенева регресія, для поліпшення точності прогнозів [7].

Методи опорних векторів (SVM) є потужними інструментами для класифікації та регресії, які можуть ефективно розділяти дані на основі складних нелінійних взаємозв'язків. Використання ядрових функцій у SVM дозволяє моделювати складні залежності між змінними, що робить цей метод особливо корисним для прогнозування ерозії ґрунту. Наприклад, для прогнозування ерозійних втрат ґрунту в умовах складного рельєфу та різноманітних кліматичних умов застосування SVM з радикальною базисною функцією може значно підвищити точність моделей.

Дерева рішень та ансамблеві методи, такі як випадковий ліс та градієнтний бустинг, також є потужними інструментами для прогнозування ерозії ґрунту. Вони можуть враховувати взаємозв'язки між численними змінними та забезпечувати високу точність прогнозів. Наприклад, випадковий ліс, що складається з 100 дерев рішень, здатен забезпечити точність прогнозу ерозії на рівні 85-90% у залежності від якості вхідних даних та параметрів моделі. Градієнтний бустинг, що використовує метод поступового додавання моделей, може додатково покращити точність прогнозів, досягаючи рівня 90-95%.

Нейронні мережі, особливо глибокі нейронні мережі та рекурентні нейронні мережі (RNN), відкривають нові можливості для прогнозування ерозії ґрунту завдяки їх здатності обробляти великі обсяги даних та враховувати складні нелінійні взаємозв'язки. Наприклад, глибока нейронна мережа з трьома прихованими шарами та 100 нейронів у кожному шарі може забезпечити високу точність прогнозів ерозії, особливо при використанні великих наборів даних з просторово-часовою інформацією. Рекурентні нейронні мережі, такі як довга короткострокова пам'ять (LSTM), можуть бути ефективними для аналізу часових рядів та прогнозування змін ерозійних процесів у часі.

Метод Random Forest продемонстрував найкращий баланс між точністю, швидкістю навчання та прогнозування порівняно з відомими результатами інших методів. Прогнозування ерозії ґрунтів за допомогою машинного навчання також відкрило нові можливості для вивчення взаємозв’язків між екологічними та економічними аспектами землекористування. Використання даних про сільськогосподарські практики та їх взаємодію з природними факторами дозволяє визначити найбільш ефективні методи управління земельними ресурсами, спрямовані на мінімізацію ерозійних процесів. Наприклад, дослідження показало, що збереження природних екосистем у поєднанні з інтеграцією органічних методів землеробства дозволяє значно зменшити ризики ерозії на довгострокову перспективу. Врахування таких факторів може суттєво змінити підхід до агрономічної діяльності та земельного планування, знижуючи негативний вплив на навколишнє середовище.

ВИСНОВКИ З ДАНОГО ДОСЛІДЖЕННЯ І ПЕРСПЕКТИВИ ПОДАЛЬШИХ РОЗВІДОК У ДАНОМУ НАПРЯМІ

Ерозія ґрунту є складним природним явищем, яке залежить від множини факторів, включаючи топографію, кліматичні умови, рослинний покрив та антропогенну діяльність. Виявлено, що традиційні методи оцінки ерозійних ризиків часто є недостатньо точними через складність інтеграції великого обсягу різнорідних даних і врахування взаємозв'язків між ними. Інтеграція методів машинного навчання в систему прогнозування ерозії ґрунту дозволяє значно підвищити точність та надійність прогнозів. Отримані результати є важливими для подальшого вдосконалення методів прогнозування ерозії ґрунтів і надають нові можливості для впровадження моделей машинного навчання в екологічну практику. Вони дозволяють розробляти адаптивні та точні стратегії боротьби з ерозією на різних етапах розвитку території — від аграрного виробництва до екологічного відновлення. Крім того, результати дослідження можуть бути використані для створення рекомендацій щодо оптимізації землекористування, планування зрошення та збереження водних ресурсів. Загальні результати досліджень підтвердили ефективність системи, її високу точність та можливість використання в реальних умовах. Архітектура системи та алгоритми машинного навчання довели свою придатність для аналізу складних взаємозв’язків між параметрами середовища. Подальше вдосконалення системи включатиме інтеграцію з геоінформаційними системами та розширення функціоналу для різних кліматичних зон.

Метод прогнозування ерозії ґрунтів з використанням машинного навчання, який поєднує емпіричні підходи та алгоритми штучного інтелекту, ґрунтується на модульній структурі. Ця структура включає етапи збору та обробки даних, навчання моделі та оцінки її ефективності. Особлива увага приділяється адаптивним методам, таким як градієнтний бустинг і нейронні мережі, що забезпечують високу точність прогнозів навіть за складних умов. Розроблена інформаційна модель враховує кліматичні, географічні та антропогенні фактори, що дозволяє налаштовувати прогнозування відповідно до специфічних умов кожного регіону.

Таким чином, концепція методу прогнозування ерозії ґрунту засобами машинного навчання базується на міждисциплінарному підході, що поєднує сучасні технології обробки даних, математичне моделювання та практичні стратегії управління земельними ресурсами. Використання таких інноваційних рішень дозволяє

значно підвищити точність прогнозування та сприяти більш ефективному використанню земельних ресурсів, що є критично важливим у сучасному світі зростаючого антропогенного тиску.

Література

1. Ерозія Ґрунтів: Причини, Види, Методи Боротьби. – EOS Data Analytics URL https://eos.com/uk/blog/eroziia-gruntiv/ (дата звернення: 03.09.2024).
2. Водна ерозія: причини, наслідки та захист грунту. EOS Data Analytics. URL: https://eos.com/uk/blog/vodna-eroziya/ (дата звернення: 23.11.2024).
3. Урожай наполовину, або Скільки коштує ерозія ґрунтів? AgroPortal. URL: https://agroportal.ua/publishing/lichnyi-vzglyad/urozhai-napolovinu-ili-skolko-stoit-eroziya-poch
4. Волощук М. Д. Ерозія грунтів Південно-західної частини України і Республіки Молдова. Вісник Харківського національного аграрного університету імені В.В. Докучаєва. Серія "Ґрунтознавство, агрохімія, землеробство, лісове господарство, екологія ґрунтів". 2016. № 1. С. 26–32. – The Library Portal of National Academy of Sciences of Ukraine (дата звернення: 06.09.2024).
5. Грабовський А. О., Ляшенко В. Р. Використання методів машинного навчання для прогнозування ерозії ґрунтів із супутникових знімків. *Journal of Soil Science and Environmental Management*, 12(3), 214-225. URL: https://doi.org/10.1016/j.jssm.2020.03.07 (дата звернення: 06.09.2024).
6. Іванов, О. В. Прогнозування ризиків ерозії ґрунтів за допомогою методів машинного навчання. *Environmental Monitoring and Assessment*, 34(2), 102-115. URL: https://doi.org/10.1016/j.emas.2021.02.11 (дата звернення: 03.09.2024).
7. Бойко П. Д., Черненко Л. О. (2020). Прогнозування ерозії ґрунтів за допомогою глибинного навчання. *Geophysical Research Letters*, 47(5), 2134-2140. URL: https://doi.org/10.1002/grl.50735 (дата звернення: 06.09.2024).
8. Івахненко О. М., Івахненко Е. Н. Ерозія ґрунту – проблема сталого аграрного землекористування. 2021. URL: https://essuir.sumdu.edu.ua/handle/123456789/85096 (дата звернення: 07.09.2024).

References

1. Eroziia Gruntiv: Prychyny, Vydy, Metody Borotby. – EOS Data Analytics URL https://eos.com/uk/blog/eroziia-gruntiv/ (data zvernennia: 03.09.2024).
2. Vodna eroziia: prychyny, naslidky ta zakhyst hruntu. EOS Data Analytics. URL: https://eos.com/uk/blog/vodna-eroziya/ (data zvernennia: 23.11.2024).
3. Urozhai napolovynu, abo Skilky koshtuie eroziia gruntiv? AgroPortal. URL: https://agroportal.ua/publishing/lichnyi-vzglyad/urozhai-napolovinu-ili-skolko-stoit-eroziya-poch
4. Voloshchuk M. D. Eroziia hruntiv Pivdenno-zakhidnoi chastyny Ukrainy i Respubliky Moldova. Visnyk Kharkivskoho natsionalnoho ahrarnoho universytetu imeni V.V. Dokuchaieva. Seriia "Hruntoznavstvo, ahrokhimiia, zemlerobstvo, lisove hospodarstvo, ekolohiia hruntiv". 2016. № 1. S. 26–32. – The Library Portal of National Academy of Sciences of Ukraine (data zvernennia: 06.09.2024).
5. Hrabovskyi A. O., Liashenko V. R. Vykorystannia metodiv mashynnoho navchannia dlia prohnozuvannia erozii gruntiv iz suputnykovykh znimkiv. Journal of Soil Science and Environmental Management, 12(3), 214-225. URL: https://doi.org/10.1016/j.jssm.2020.03.07 (data zvernennia: 06.09.2024).
6. Ivanov, O. V. Prohnozuvannia ryzykiv erozii gruntiv za dopomohoiu metodiv mashynnoho navchannia. Environmental Monitoring and Assessment, 34(2), 102-115. URL: https://doi.org/10.1016/j.emas.2021.02.11 (data zvernennia: 03.09.2024).
7. Boiko P. D., Chernenko L. O. (2020). Prohnozuvannia erozii gruntiv za dopomohoiu hlybynnoho navchannia. Geophysical Research Letters, 47(5), 2134-2140. URL: https://doi.org/10.1002/grl.50735 (data zvernennia: 06.09.2024).
8. Ivakhnenko O. M., Ivakhnenko E. N. Eroziia gruntu – problema staloho ahrarnoho zemlekorystuvannia. 2021. URL: https://essuir.sumdu.edu.ua/handle/123456789/85096 (data zvernennia: 07.09.2024).